Biostat 200C Homework 2

Due Apr 30 @ 11:59PM

Q1. Beta-Binomial

Let Y_i be the number of successes in n_i trials with

$$Y_i \sim \text{Bin}(n_i, \pi_i),$$

where the probabilities π_i have a Beta distribution

$$\pi \sim \text{Be}(\alpha, \beta)$$

with density function

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1], \alpha > 0, \beta > 0.$$

Q1.1

Find the mean and variance of π.

Answer Mean:

$$\begin{aligned}
E(\pi_i) &= \int \pi_i * f(\pi_i) d\pi_i \\
&= \int \pi_i * \pi_i^{\alpha-1}(1-\pi_i)^{\beta-1}/B(\alpha,\beta) d\pi_i \\
&= B(\alpha,\beta)^{-1} \int \pi_i^{(\alpha+1)-1}(1-\pi_i)^{\beta-1} d\pi_i \\
&= B(\alpha+1,\beta) * B(\alpha,\beta)^{-1} \int B(\alpha+1,\beta)^{-1} * \pi_i^{(\alpha+1)-1}(1-\pi_i)^{\beta-1} d\pi_i \\
&= B(\alpha+1,\beta) * B(\alpha,\beta)^{-1} * 1 \\
&= \Gamma(\alpha+1)\Gamma(\beta)/\Gamma(\alpha+1+\beta)/(\Gamma(\alpha)\Gamma(\beta)) * \Gamma(\alpha+\beta) \\
&= \alpha/(\alpha+\beta) \\
&= \theta
\end{aligned}$$

Variance: Define $\frac{1}{\alpha+\beta+1} = \phi$

$$
\begin{aligned}
E(\pi_i^2) &= \int \pi_i^2 * f(\pi_i) d\pi_i \\
&= \int \pi_i^2 * \pi_i^{\alpha-1}(1-\pi_i)^{\beta-1}/B(\alpha,\beta) d\pi_i \\
&= B(\alpha,\beta)^{-1} \int \pi_i^{(\alpha+2)-1}(1-\pi_i)^{\beta-1} d\pi_i \\
&= B(\alpha+2,\beta) * B(\alpha,\beta)^{-1} \int B(\alpha+1,\beta)^{-1} * \pi_i^{(\alpha+2)-1}(1-\pi_i)^{\beta-1} d\pi_i \\
&= B(\alpha+2,\beta) * B(\alpha,\beta)^{-1} * 1 \\
&= \Gamma(\alpha+2)\Gamma(\beta)/\Gamma(\alpha+2+\beta)/(\Gamma(\alpha)\Gamma(\beta)) * \Gamma(\alpha+\beta) \\
&= \alpha * (\alpha+1)/(\alpha+1+\beta) * (\alpha+\beta) \\
&= \theta(\alpha+1)/(\alpha+1+\beta)
\end{aligned}
$$

Then we can obtain $Var(\pi_i)$

$$
\begin{aligned}
Var(\pi_i) &= E(\pi_i^2) - E(\pi_i)^2 \\
&= ((\alpha+1)\alpha(\alpha+\beta) - \alpha^2(\alpha+\beta+1))/(\alpha+\beta+1)(\alpha+\beta)^2 \\
&= (\alpha\beta)/(\alpha+\beta)^2/(\alpha+1+\beta) \\
&= \theta(1-\theta)/(\alpha+\beta+1) = \phi\theta(1-\theta)
\end{aligned}
$$

Q1.2

Find the mean and variance of Y_i and show that the variance of Y_i is always larger than or equal to that of a Binomial random variable with the same batch size and mean.

Answer:

$$
\begin{aligned}
Var(Y_i) &= E_{\pi_i}(Var(Y_i|\pi_i)) + Var_{\pi_i}(E(Y_i|\pi_i)) \\
&= E_{\pi_i}(n_i * \pi_i * (1-\pi_i)) + Var_{\pi_i}(\pi_i * n_i) \\
&= n_i * (E(\pi_i) - E(\pi_i^2)) + n_i^2 * \phi\theta(1-\theta) \\
&= n_i * (\theta - \theta(\alpha+1)/(\alpha+1+\beta)) + n_i^2 * \phi\theta(1-\theta) \\
&= n_i * (\theta(1-(\alpha+1)/(\alpha+1+\beta))) + n_i^2 * \phi\theta(1-\theta) \\
&= n_i * (\theta * \beta/(\alpha+1+\beta)) + n_i^2 * \phi\theta(1-\theta) \\
&= n_i * (\theta * (1-\theta)(1-\phi)) + n_i^2 * \phi\theta(1-\theta) \\
&= n_i\theta(1-\theta)[1+(n_i-1)\phi]
\end{aligned}
$$

Then we know that $Var(Y_i) = n_i\theta(1-\theta)[1+(n_i-1)\phi]$ so that $Var(Y_i)$ is larger than the Binomial variance (unless $n_i = 1$ or $\phi = 0$).

Q2. Motivation for quasi-binomial

Verify that the log-likilihood ℓ_i of a binomial proportion Y_i, where $m_iY_i \sim \mathrm{Bin}(m_i, p_i)$, satisfies

$$\begin{aligned}
\mathbb{E}\frac{\partial \ell_i}{\partial \mu_i} &= 0 \\
\mathrm{Var}\,\frac{\partial \ell_i}{\partial \mu_i} &= \frac{1}{\phi V(\mu_i)} \\
\mathbb{E}\frac{\partial^2 \ell_i}{\partial \mu_i^2} &= -\frac{1}{\phi V(\mu_i)},
\end{aligned}$$

with $\phi = 1$, $\mu_i = p_i$, and $V(\mu_i) = p_i(1-p_i)/m_i$. Therefore the U_i in quasi-binomial method mimics the behavior of a binomial model.

Answer: (1) As for $\mathbb{E}\frac{\partial \ell_i}{\partial \mu_i}$

$$\begin{aligned}
\ell_i(\boldsymbol{\beta}) =& m_iy_i \log p_i + (m_i - m_iy_i)\log(1-p_i) + \log\binom{m_i}{m_iy_i} \\
\ell_i(\boldsymbol{\beta}) =& m_iy_i \log \mu_i + (m_i - m_iy_i)\log(1-\mu_i) + \log\binom{m_i}{m_iy_i} \\
\implies \frac{\delta \ell_i}{\delta \mu_i} =& \frac{m_iy_i}{\mu_i} - \frac{m_i - m_iy_i}{1-\mu_i} \\
=& \frac{m_iy_i * (1-\mu_i) - (m_i - m_iy_i) * \mu_i}{\mu_i * (1-\mu_i)} \\
=& \frac{m_iy_i - m_i\mu_i}{\mu_i * (1-\mu_i)} \\
\implies E(\frac{\delta \ell_i}{\delta \mu_i}) =& E(\frac{m_iy_i - m_i\mu_i}{\mu_i * (1-\mu_i)}) \\
=& \frac{E(m_iy_i - m_i\mu_i)}{E(\mu_i * (1-\mu_i))} \\
=& \frac{m_iE(y_i) - E(m_i\mu_i)}{E(\mu_i * (1-\mu_i))} \\
=& \frac{m_i\mu_i - m_i\mu_i}{E(\mu_i * (1-\mu_i))} \\
=& 0
\end{aligned}$$

(2) As for $\mathrm{Var}\,\frac{\partial \ell_i}{\partial \mu_i}$

$$
\begin{aligned}
\operatorname{Var}\frac{\partial \ell_i}{\partial \mu_i} &= E((\frac{\partial \ell_i}{\partial \mu_i})^2) - E(\pi_i)^2 \\
&= E((\frac{\partial \ell_i}{\partial \mu_i})^2) \\
&= E((\frac{y_i - m_i\mu_i}{\mu_i * (1-\mu_i)})^2), \text{ according to (1)} \\
&= E(\frac{y_i^2 - 2y_i m_i \mu_i + (m_i\mu_i)^2}{\mu_i^2 * (1-\mu_i)^2}) \\
&= \frac{E(y_i^2) - 2m_i\mu_i E(2y_i) + (m_i\mu_i)^2}{\mu_i^2 * (1-\mu_i)^2} \\
&= \frac{var(y_i) - E(y_i)^2 - 2m_i\mu_i * m_i\mu_i + (m_i\mu_i)^2}{\mu_i^2 * (1-\mu_i)^2} \\
&= \frac{m_i p_i(1-p_i) + (m_i\mu_i)^2 - 2m_i\mu_i * m_i\mu_i + (m_i\mu_i)^2}{\mu_i^2 * (1-\mu_i)^2} \\
&= \frac{m_i p_i(1-p_i)}{\mu_i^2 * (1-\mu_i)^2} \\
&= \frac{1}{\frac{\mu_i(1-\mu_i)}{m_i}} \\
&= \frac{1}{\phi V(\mu_i)}, \phi = 1
\end{aligned}
$$

(3) As for $\mathbb{E}\frac{\partial^2 \ell_i}{\partial \mu_i^2}$

From (1) we know that,

$$
\begin{aligned}
\frac{\delta \ell_i}{\delta \mu_i} =& \frac{m_i y_i}{\mu_i} - \frac{m_i - y_i}{1-\mu_i} \\
\implies \frac{\partial^2 \ell_i}{\partial \mu_i^2} =& \frac{-m_i y_i}{\mu_i^2} - \frac{m_i - m_i y_i}{(1-\mu_i)^2} \\
=& \frac{-m_i y_i (1-\mu_i)^2 - \mu_i^2 (m_i - m_i y_i)}{\mu_i^2 (1-\mu_i)^2} \\
=& \frac{-m_i y_i (1 - 2\mu_i + \mu_i^2) - \mu_i^2 m_i + \mu_i^2 y_i}{\mu_i^2 (1-\mu_i)^2} \\
=& \frac{-m_i y_i (1-2\mu_i) - \mu_i^2 m_i}{\mu_i^2 (1-\mu_i)^2} \\
\implies E(\frac{\partial^2 \ell_i}{\partial \mu_i^2}) =& E(\frac{-m_i y_i (1-2\mu_i) - \mu_i^2 m_i}{\mu_i^2 (1-\mu_i)^2}) \\
=& \frac{E(-m_i y_i (1-2\mu_i) - \mu_i^2 m_i)}{E(\mu_i^2 (1-\mu_i)^2)} \\
=& \frac{(2\mu_i - 1) m_i E(y_i) - \mu_i^2 m_i}{\mu_i^2 (1-\mu_i)^2} \\
=& \frac{(2\mu_i - 1)\mu_i m_i - \mu_i^2 m_i}{\mu_i^2 (1-\mu_i)^2} \\
=& \frac{(\mu_i - 1)\mu_i m_i}{\mu_i^2 (1-\mu_i)^2} \\
=& -\frac{m_i}{\mu_i (1-\mu_i)} \\
=& -\frac{1}{\frac{p_i(1-p_i)}{m_i}} \\
=& -\frac{1}{\phi V(\mu_i)}, \phi = 1
\end{aligned}
$$

Q3. Concavity of Poisson regression log-likelihood

Let $Y_1, \dots, Y_n$ be independent random variables with $Y_i \sim \text{Poisson}(\mu_i)$ and $\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $i = 1, \dots, n$.

Q3.1

Write down the log-likelihood function.

Answer:

$$
\begin{aligned}
\ell(\boldsymbol{\beta}) =& \sum_i (y_i \log \mu_i - \mu_i - \log y_i!) \\
=& \sum_{i=1}^n (y_i \cdot \mathbf{x}_i^T \boldsymbol{\beta} - e^{\mathbf{x}_i^T \boldsymbol{\beta}} - \log y_i!)
\end{aligned}
$$

Q3.2

Derive the gradient vector and Hessian matrix of the log-likelihood function with respect to the regression coefficients $\boldsymbol{\beta}$.

Answer:

(1) As for the gradient vector,

$$\begin{aligned}\nabla f(\beta) &= \sum_{i=1}^{n} \frac{\delta \ell_i}{\delta \beta} \\ &= \sum_{i=1}^{n} y_i \cdot \mathbf{x}_i - e^{\mathbf{x}_i^T \beta} \mathbf{x}_i\end{aligned}$$

(2) As for the Hessian matrix,

$$\begin{aligned}\mathbf{H}(\beta) &= \sum_{i=1}^{n} \frac{\delta^2 \ell_i}{\delta \beta^2} \\ &= \sum_{i=1}^{n} (y_i \cdot \mathbf{x}_i - e^{\mathbf{x}_i^T \beta} \mathbf{x}_i)' \\ &= \sum_{i=1}^{n} -e^{\mathbf{x}_i^T \beta} \mathbf{x}_i \mathbf{x}_i^T \\ &= -\mathbf{x}^T e^{\mathbf{x}^T \beta} \mathbf{x}\end{aligned}$$

Q3.3

Show that the log-likelihood function of the log-linear model is a concave function in regression coefficients $\boldsymbol{\beta}$. (Hint: show that the negative Hessian is a positive semidefinite matrix.)

Answer:

According to the Q3.2, we know that $\mathbf{H}(\beta) = -\mathbf{x}^T e^{\mathbf{x}^T \beta} \mathbf{x}$. Then we can have an arbitrary vector $\mathbf{v}$,

$$\begin{aligned}\mathbf{v}^T(-\mathbf{H}(\beta))\mathbf{v} &= \mathbf{v}^T \mathbf{x}^T e^{\mathbf{x}^T \beta} \mathbf{x} \mathbf{v} \\ &= e^{\mathbf{x}^T \beta} \mathbf{v}^T \mathbf{x}^T \mathbf{x} \mathbf{v}\end{aligned}$$

Since we know that $\boldsymbol{\mu} = e^{\mathbf{x}^T \beta} > 0$ and $\mathbf{v}^T \mathbf{x}^T \mathbf{x} \mathbf{v} = (\mathbf{v}\mathbf{x})^T \mathbf{x} \mathbf{v} \geq 0$, then we know $\mathbf{v}^T(-\mathbf{H}(\beta))\mathbf{v} \geq 0$ for all $\mathbf{v}$, which indicates that $-\mathbf{H}(\beta)$ is a positive semi-definite matrix. Thus, the log-linear model is a concave function in regression coefficients.

Q3.4

Show that for the fitted values $\widehat{\mu}_i$ from maximum likelihood estimates

$$\sum_i \widehat{\mu}_i = \sum_i y_i.$$

Therefore the deviance reduces to

$$D = 2 \sum_i y_i \log \frac{y_i}{\widehat{\mu}_i}.$$

Answer:

$$
\begin{aligned}
\ell(\mu_i) &= \sum_i (y_i \log \mu_i - \mu_i - \log y_i!) \\
\implies \frac{\delta \ell_i}{\delta \mu_i} &= \sum_i \frac{y_i}{\mu_i} - 1
\end{aligned}
$$

Then we set $\frac{\delta \ell_i}{\delta \mu_i} = 0$ to obtain maximum likelihood estimates for μ_i,

$$
\begin{aligned}
\frac{\delta \ell_i}{\delta \widehat{\mu}_i} =& 0 \\
\sum_i \frac{y_i}{\widehat{\mu}_i} - 1 =& 0 \\
\sum_i \frac{y_i}{\widehat{\mu}_i} =& 1 \\
\sum_i y_i =& \sum_i \widehat{\mu}_i
\end{aligned}
$$

Then we plug in this results into the deviance,

$$
\begin{aligned}
D &= 2\sum_i [y_i \log(y_i) - y_i] - 2\sum_i [y_i \log(\widehat{\mu}_i) - \widehat{\mu}_i] \\
&= 2\sum_i [y_i \log(y_i/\widehat{\mu}_i) - (y_i - \widehat{\mu}_i)] \\
&= 2\sum_i [y_i \log(y_i/\widehat{\mu}_i)] - (\sum_i y_i - \sum_i \widehat{\mu}_i) \\
&= 2\sum_i y_i \log(y_i/\widehat{\mu}_i)
\end{aligned}
$$

Q4. Odds ratios

Consider a 2×2 contingency table from a prospective study in which people who were or were not exposed to some pollutant are followed up and, after several years, categorized according to the presense or absence of a disease. Following table shows the probabilities for each cell. The odds of disease for either exposure group is $O_i = \pi_i/(1 - \pi_i)$, for $i = 1, 2$, and so the odds ratio is

$$
\phi = \frac{O_1}{O_2} = \frac{\pi_1(1 - \pi_2)}{\pi_2(1 - \pi_1)}
$$

is a measure of the relative likelihood of disease for the exposed and not exposed groups.

	Diseased	Not diseased
Exposed	π_1	$1 - \pi_1$
Not exposed	π_2	$1 - \pi_2$

Q4.1

For the simple logistic model

$$
\pi_i = \frac{e^{\beta_i}}{1 + e^{\beta_i}},
$$

show that if there is no difference between the exposed and not exposed groups (i.e., $\beta_1 = \beta_2$), then $\phi = 1$.

Answer: Given that$\beta_1 = \beta_2$, we know that,

$$\begin{aligned}\pi_1 &= \frac{e^{\beta_1}}{1+e^{\beta_1}} \\ &= \frac{e^{\beta_2}}{1+e^{\beta_2}} \\ &= \pi_2\end{aligned}$$

Then plug it in the equation of odds ratio,

$$\begin{aligned}\phi &= \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} \\ &= \frac{\pi_1(1-\pi_1)}{\pi_1(1-\pi_1)} \\ &= 1\end{aligned}$$

Q4.2

Consider J 2×2 tables, one for each level x_j of a factor, such as age group, with $j = 1, \ldots, J$. For the logistic model

$$\pi_{ij} = \frac{e^{\alpha_i + \beta_i x_j}}{1 + e^{\alpha_i + \beta_i x_j}}, \quad i = 1, 2, \quad j = 1, \ldots, J.$$

Show that $\log \phi$ is constant over all tables if $\beta_1 = \beta_2$.

Answer: Define $\beta = \beta_1 = \beta_2$, then we can have,

$$\begin{aligned}\pi_{1j} &= \frac{e^{\alpha_1 + \beta x_j}}{1 + e^{\alpha_1 + \beta x_j}}, \quad j = 1, \ldots, J. \\ \pi_{2j} &= \frac{e^{\alpha_2 + \beta x_j}}{1 + e^{\alpha_2 + \beta x_j}}, \quad j = 1, \ldots, J.\end{aligned}$$

Then we plug it in the equation of log odds ratio,

$$\begin{aligned}\log \phi_j &= \log \frac{\pi_{1j}(1-\pi_{2j})}{\pi_{2j}(1-\pi_{1j})} \\ &= \log \frac{\frac{e^{\alpha_1+\beta x_j}}{1+e^{\alpha_1+\beta x_j}}(1 - \frac{e^{\alpha_2+\beta x_j}}{1+e^{\alpha_2+\beta x_j}})}{\frac{e^{\alpha_2+\beta x_j}}{1+e^{\alpha_2+\beta x_j}}(1 - \frac{e^{\alpha_1+\beta x_j}}{1+e^{\alpha_1+\beta x_j}})} \\ &= \log \frac{\frac{e^{\alpha_1+\beta x_j}}{1+e^{\alpha_1+\beta x_j}} \frac{1}{1+e^{\alpha_2+\beta x_j}}}{\frac{e^{\alpha_2+\beta x_j}}{1+e^{\alpha_2+\beta x_j}} \frac{1}{1+e^{\alpha_1+\beta x_j}}} \\ &= \log \frac{e^{\alpha_1+\beta x_j}}{e^{\alpha_2+\beta x_j}} \\ &= \log e^{\alpha_1 - \alpha_2} \\ &= \alpha_1 - \alpha_2\end{aligned}$$

Then we know that $\log \phi$ is constant $\alpha_1 - \alpha_2$ when $\beta_1 = \beta_2$.